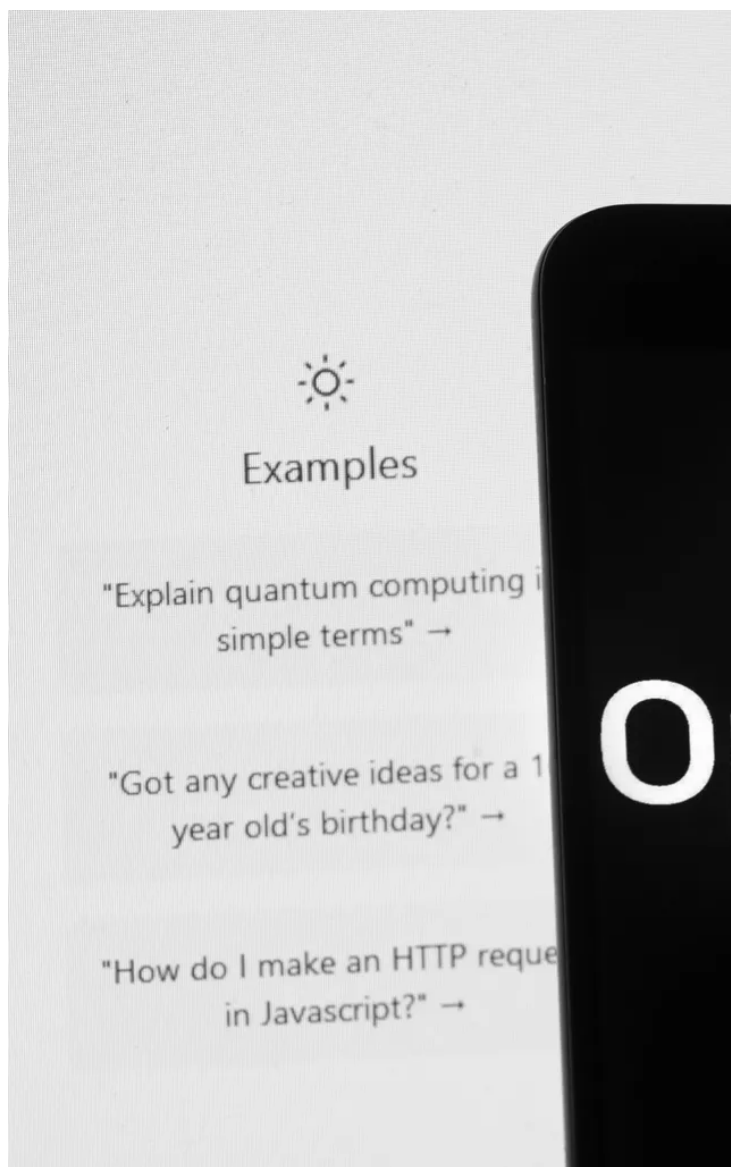


The chatbot era: Better or worse off?

March 31, 2023 at 1:08 pm



The OpenAI logo is seen on a mobile phone in front of a computer screen that displays the ChatGPT home screen, Friday, March 17, 2023, in Boston. (Michael Dwyer / The Associated Press)

By [Jevin West](#)

Special to The Times

Are we better off because of penicillin? Yes. The internet? Probably. Social media? Probably not.

So what about chatbots? The chatbot craze has captured the world's attention, and [massive piles of money](#). Chatbots are software programs that use artificial intelligence to process and simulate conversations with humans. Will they improve human experience and longevity, peace and prosperity, environmental health, productivity or social well-being?

From my perspective, as a researcher who studies misinformation and its effects on society, chatbots will be vectors of propaganda, they will make it harder to discern truth, and they will further erode trust in our institutions. I see two main reasons for this: They are bullshitters at scale, and they are difficult, if not impossible, to reverse engineer.

In 2017, the Federal Communications Commission invited the public to submit comments regarding net neutrality. More than 20 million were submitted. The problem is that a [large number of these](#) were from fake identities. Grandma's handwritten letter had no chance of being heard.

This was pre-ChatGPT, the celebrated chatbot released in November by Microsoft-backed OpenAI. Imagine how easy it will be to flood public commentary with a bot that has, [according to some](#), passed the Turing test.

Even in situations with less nefarious intent, chatbots are purveyors of misinformation. I teach a class on bullshit — an act, often with full confidence, intended to persuade, with no allegiance to truth. My colleague and co-instructor of the class, Carl Bergstrom, asked Galactica, Meta's large language equivalent, to describe Brandolini's BS asymmetry principle: "the amount of energy needed to refute BS is an order of magnitude bigger than that needed to produce it."

Galactica's answer: "a theory in economics [Not True] proposed by Gianni Brandolini [Not True], a professor at the University of Padua [Not True], which states that 'the

smaller economic unit, the greater its efficiency [Not True] ...” The falsehoods continue for another two paragraphs, but I think you get the point. It was BS-ing the BS principle. This chatbot, like others, answers right and wrong with the confidence of a car salesman.

Qualifiers of confidence (e.g., “I am pretty sure”) are important human signals that bots have not mastered. My feature request for ChatGPT-10 is a confidence meter. Currently, when I ask ChatGPT about its confidence in an answer, it repeatedly replies, “I do not experience confidence or doubt in the way that humans do.” I wish it did.

My colleagues and friends who are more bullish on chatbots say, “But, Jevin, look how many things they get right.” It is amazing — magical even. But just like magic shows, people want to believe. When the magician flubs, we give them a pass. When they get it right, we cheer with joy. The problem with chatbots is that their show won’t end any time soon.

True, they get a lot right, but even if they only get 10% wrong, that adds up over OpenAI’s [100 million monthly users](#), especially when those answers involve someone’s health and safety.

It is not just the error rate that is problematic. The bigger problem is how difficult, if not impossible, it is to reverse engineer the wrong answers.

When Bing’s chatbot [went off the rails](#) with a New York Times reporter, engineers couldn’t easily just say, “Oh, whoops, I set that parameter wrong; let me fix that.” Large language models don’t work that way. Their billions of parameters are trained on petabytes of data. These are not programs with single-line instructions that can be debugged easily. When things go wrong, some techniques exist, like [reinforcement learning](#), but for the most part, the techniques are Band-Aids trying to patch a gaping methodological wound.

Reverse engineering why a chatbot would ask a human to leave their spouse is one challenge. Another challenge is trying to reverse engineer where an answer even comes from. As a teacher and researcher, I train my students to support their claims with references, and when presented with claims, check the sources. This is how to debunk false claims and verify true claims. Many of the current bots don’t do this. These autocomplete machines spit out an answer. That’s it. Most don’t provide the sources on which the answers originated and when they are asked, many [make up references](#).

Summaries devoid of sources will glue eyeballs to search engines and their advertisements longer, but it will not help fact-checkers debunk claims or provide financial resources to the originators of content on which these bots are trained.

Other problems exist — the environmental cost for training these large language models, the impact on human connection, the dangerous feedback loops from bots training on their own output, etc. — but it is the impact these bots could have on the health and integrity of our information environments that concerns me most.

I hope the future proves me wrong, demonstrates the myriad ways chatbots reduce the spread of harmful misinformation and disinformation, and shows why the world is better off. I am “pretty sure,” though, that bots are not the antidote we have been looking for.

Postscript: What about your professional and personal lives? Will your world be better or worse off with these bots? Leave comments below (human generated answers only). If there is enough interest, I will help organize a public seminar on this topic at the University of Washington.

Jevin West is an associate professor at the University of Washington's Information School and co-founder of the Center for an Informed Public.

 [View 13 Comments](#)